

EDUCATION

Berkeley, CA	University of California at Berkeley
Aug. 2017 - May 2021	Bachelor of Arts in Data Science; Domain Emphasis in Linguistics; Minor in Computer Science National Merit Scholar; Valedictorian of Lamar High School GRE: Verbal: 168, Math: 164, Overall: 332/340
Coursework:	<div>Algorithms</div> <div>Data Structures</div> <div>Data Engineering</div> <div>Machine Learning and Data Analytics</div> <div>Artificial Intelligence (AI)</div> <div>Linux System Administration</div> <div>Natural Language Processing (NLP)</div> <div>Data Mining and Analytics</div> <div>Computer Architecture</div> <div>Designing Information Devices and Systems</div> <div>Discrete Math and Probability Theory</div> <div>Linear Algebra and Differential Equations</div>

EXPERIENCE

Remote	Smarsh - Software Engineer
Starting in June 2021	<ul style="list-style-type: none">Building new features for Smarsh's production Conduct Surveillance application and solving production problems escalated from the support team.
Berkeley, CA	University of California at Berkeley, Cohen Group - Research Intern
Feb. 2019 - Aug. 2019	<ul style="list-style-type: none">Launched efforts to replace a computationally expensive WRF-CHEM climate model run with machine learning model predictions.Implemented preparations to produce higher resolution measurements of NO2 levels than measured by NASA's Aura satellite.Aligned 47K WRF-CHEM pixels with Aura's OMI pixels to enable better data correlation.
Austin, TX	University of Texas at Austin - Summer Student Researcher
Jun. 2016 - Aug. 2016	<ul style="list-style-type: none">Optimized catalysts for improved efficiency in energy production.Created molecular visualization movies of catalyst binding to communicate research.Computationally simulated carbon monoxide binding on alloys of gold-palladium nanoparticles during oxidation-reduction reactions.

PROJECTS

More on my website: <https://dannysiu.com>

Python (PyTorch, NumPy)	POS-Tagger Model <ul style="list-style-type: none">Implemented and trained an LSTM for POS-tagging by using mini-batch stochastic gradient descent on pre-trained GloVe embeddings.Improved model accuracy by implementing a bi-directional LSTM and dropout layer, and loading the 500,000 most common 300-dimensional word embeddings.
Python (Scikit-Learn, Matplotlib, Pandas, Seaborn, NumPy)	Spam Email Classifier <ul style="list-style-type: none">Performed exploratory data analysis on textual data to perform feature selection.Implemented feature engineering for a logistic regression model and utilized carefully selected spam keywords to fit my spam classifier model.
Python (Scikit-Learn, SciPy, NumPy)	Movie Review Sentiment Classifier <ul style="list-style-type: none">Created a binary sentiment classifier for movie reviews using featurization of bag-of-words, bigrams, trigrams, sentiment dictionaries, and neutral word removal.
Java	PokeMan Game <ul style="list-style-type: none">Developed a multi-player game featuring Pokemon characters in randomized worlds.Implemented game saving using Git, menus, and a live heads-up display (HUD).

SKILLS

Programming:	Python, Java, SQL, PostgreSQL, Linux, Git, C, PyTorch, HTML, CSS, Relational DBMS, Bash scripting, Jupyter notebooks
Data Science:	Data modeling (Classification, Prediction, Clustering, Gradient descent), Data visualization, Machine learning, Feature engineering, Data analysis, Data mining
NLP:	Neural networks (RNNs, CNNs), LSTMs, Hidden Markov Models, Sentiment analysis, Dependency parsing, Text embeddings (Word2vec, GloVe, FastText), One-hot encoding